Appendix C

Additional Information Requested by the Panel

[This Page Intentionally Left Blank]

Appendix C1

ESAC Statement on the Validity of *In Vitro* Pyrogen Tests

[This Page Intentionally Left Blank]

EUROPEAN COMMISSION
DIRECTORATE GENERAL JRC
JOINT RESEARCH CENTRE
Institute for Health and Consumer Protection
European Centre for the Validation of Alternative Methods (ECVAM)

STATEMENT ON THE VALIDITY OF IN-VITRO PYROGEN TESTS

At its 24th meeting, held on 20-21 March 2006 at the European Centre for the validation of alternative methods (ECVAM), Ispra, Italy, the non-Commission members of the ECVAM Scientific Advisory Committee (ESAC)[1] unanimously endorsed the following statement:

Following a review of scientific reports and peer reviewed publications on the following range of in-vitro pyrogen tests:

1. Human Whole Blood IL-1,
2. Human Whole Blood IL-6,
3. PBMC IL-6,
4. MM6 IL-6, and
5. Human Cryopreserved Whole Blood IL-1,

it is concluded that these tests have been scientifically validated for the detection of pyrogenicity mediated by Gram-negative endotoxins, and quantification of this pyrogen, in materials currently evaluated and characterized by rabbit pyrogen tests.

These methods have the potential to satisfy regulatory requirements for the detection and quantification of these pyrogens in these materials subject to product-specific validation.

The test methods have the capacity of detecting pyrogenicity produced by a wider range of pyrogens, but the evidence compiled for, and considered within this peer review and validation process, is not sufficient to state that full scientific validation of this wider domain of applicability has been demonstrated and confirmed.

Thus, the above test methods can currently be considered as full replacements for the evaluation of materials or products where the objective is to identify and evaluate pyrogenicity produced by Gram-negative endotoxins, but not for other pyrogens.

This endorsement takes account of the dossiers prepared for peer review; the views of independent experts who evaluated the dossiers against defined validation criteria; supplementary submissions made by the Management Team; and the considered view of the Peer Review Panel appointed to oversee the process.

Thomas Hartung
Head of Unit
ECVAM
Institute for Health & Consumer Protection
Joint Research Centre
European Commission
Ispra

21 March 2006

1. The ESAC was established by the European Commission, and is composed of nominees from the EU Members States, industry, academia and animal welfare, together with representatives of the relevant Commission services.

   This statement was endorsed by the following Members of the ESAC:

   Prof Helmut Tritthart (Austria)
   Dr Dagmar Jírová (Czech Republic)
   Prof Elisabeth Knudsen (Denmark)
   Dr Timo Ylikomi (Finland)
   Prof André Guillouzo (France)
   Dr Manfred Liebsch (Germany)
   Dr Efstathios Nikolaidis (Greece)
   Dr Katalin Horvath (Hungary)
   Prof Michael Ryan (Ireland)
   Dr Annalaura Stammati (Italy)
   Dr Mykolas Maurica (Lithuania)
   Prof Eric Tschirhart (Luxembourg)
   Dr Jan van der Valk (The Netherlands)
   Dr Dariusz Sladowski (Poland)
   Prof Milan Pogačnik (Slovenia)
   Dr Argelia Castaño (Spain)
   Dr Patric Amcoff (Sweden)
   Dr Jon Richmond (UK)
   Dr Odile de Silva (COLIPA)
   Dr Julia Fentem (ECETOC)
   Dr Nathalie Alépée (EFPIA)
   Prof Robert Combes (ESTIV)
   Dr Maggy Jennings (Eurogroup for Animal Welfare)
   Mr Roman Kolar (Eurogroup for Animal Welfare)

   The following Commission Services and Observer Organisations were involved in the consultation process, but not in the endorsement process itself.

   Mr Thomas Hartung (ECVAM; chairman)
   Mr Jens Linge (ECVAM; ESAC secretary)
   Mr Juan Riego Sintes (ECB)
   Ms Beatrice Lucaroni (DG Research, Unit F.5)
   Mr Sylvain Bintein (DG Environment, Unit C.3)
   Mr Sigfried Breier (DG Enterprise, Unit F.3)
   Prof Dr Constantin Mircioiu (Romania)
   Dr William Stokes (NICEATM, USA)
   Prof Dr Vera Rogiers (ECOPA)

Annex

The novel pyrogen tests are based on the human fever reaction. Monocytoid cells, either primary from human blood or as propagated cell lines, detect pyrogens of different chemical nature and respond by the release of inflammatory mediators such as cytokines. Since lipopolysaccharides from Gram-negative bacteria are the only type of proven pyrogen, for which an International reference material is available, the tests were standardised to detect the presence of significantly less than 0.5 Endotoxin Units of this preparation, which is considered to be the threshold level for fever induction in the most sensitive rabbit species according to pharmacopoeia test procedures.
The five tests which were sufficiently reproducible and exceeded the rabbit test with regard to sensitivity and specificity for the detection of lipopolysaccharide spiked samples, differ with regard to cell source and preparation, cryopreservation and cytokine measured. The tests have been described elsewhere (1-4). The concept of the validation study (5) and the international validation studies are available (6-7).

1. Poole, S., Thorpe, R., Meager, A., Hubbard, A.R., Gearing, A.J. (1988) Detection of pyrogen by cytokine release. Lancet 8577, 130.

2. Taktak, Y.S., Selkirk, S., Bristow, A.F., Carpenter, A., Ball, C., Rafferty, B., Poole, S. (1991) Assay of pyrogens by interleukin-6 release from monocytic cell lines. J. Pharm. Pharmacol. 43, 578.

3. Hartung, T., Wendel, A. (1996) Detection of pyrogens using human whole blood. In Vitro Toxicol. 9, 353.

4. Schindler S, Asmus S, von Aulock S, Wendel A, Hartung T and Fennrich S. (2004) Cryopreservation of human whole blood for pyrogenicity testing. J. Immunol. Meth. 294, 89-100.

5. Hartung, T., Aaberge, I., Berthold, S., Carlin, G., Charton, E., Coecke, S., Fennrich, S., Fischer, M., Gommer, M., Halder, M., Haslov, K., Jahnke, M., Montag-Lessing, T., Poole, S., Schechtman, L., Wendel, A., Werner-Felmayer, G. (2001) Novel pyrogen tests based on the human fever reaction. The report and recommendations of ECVAM Workshop 43. European Centre for the Validation of Alternative Methods. Altern. Lab. Anim. 29, 99.

6. Hoffmann S, Peterbauer A, Schindler S, Fennrich S, Poole S. Mistry Y, Montag-Lessing T, Spreitzer I, Loschner B, vam Aalderen M, Bos R, Gommer M, Nibbeling R, Werner-Felmayer G, Loitzl P, Jungi T, Brcic M, Brugger P, Frey E, Bowe G, Casado J, Coecke S, de Lange J, Mogster B, Naess LM, Aaberge IS, Wendel A and Hartung T. (2005) International validation of novel pyrogen tests based on the human fever reaction. J. Immunol. Meth. 298, 161-173.

7. Schindler S, Spreitzer I, Loschner, Hoffmann S, Hennes K, Halder M, Brügger P, Frey E, Hartung T and Montag T. (2006) International validation of pyrogen tests based on cryopreserved human primary blood cells. J. Immunol. Meth. 316, 42-51.

[This Page Intentionally Left Blank]

Appendix C2

Press Release: "Fewer Tests on Animals and Safer Drugs: New EU Tests Save 200,000 Rabbits per Year"

[This Page Intentionally Left Blank]

IP

Brussels, 12 May 2003

Fewer tests on animals and safer drugs: new EU tests save 200,000 rabbits per year

New, groundbreaking methods of drug testing to replace animals with safe alternatives, saving up to 200,000 rabbits per year, were unveiled today in Brussels by European Research Commissioner Philippe Busquin. The set of six tests detects potential fever-causing agents (pyrogens) in drugs, by using human blood cells instead of rabbits. The new tests have been developed by a EU-supported research team, involving national control laboratories, test developers, and companies. The tests are being validated by the Commission. They are already being used in over 200 laboratories across the world. Thanks to these alternative methods rabbits will no longer be needed to test the presence of pyrogens in parenteral (non oral) drugs.

"The use of animals to test drugs is unfortunately necessary to safeguard human health," said European Research Commissioner Philippe Busquin. "But we can reduce, replace and refine animal testing, with EU-sponsored research leading the way at world level. The EU's validation of these new testing methods will encourage their broad take-up by industry, ensure drug safety and quality, and reduce the use of animal research. This is an example of the European Research Area in action, developing an environment in which scientific results can be rapidly exploited and transformed into products and processes that improve quality of life, increase competitiveness and benefit animal welfare."

The safety and potency of commercially available medicines and vaccines must be guaranteed. Innovative research, funded and validated by the Commission, aims to replace existing animal-based test methods for fever-causing agents (pyrogens) in parenteral drugs with a new generation of in vitro tests that are more accurate, quicker and more cost-effective.

Blood cells replace rabbits

Understanding of human immunology has advanced rapidly in the past 20 years. Work on human fever reaction and development of test systems for fever mediator molecules, combined with improved cell biology techniques, now enables the innovative use of human cells as biosensors for pyrogens (fever-causing agents). The EU study[1] set out to compare and harmonise six in vitro assays to develop a "state-of-the-art" method for inclusion into the European Pharmacopoeia - which sets the requirements for the quality control of drugs in Europe - thus improving consumer safety.

The EU role

[1] Cell factory project: *Comparison and validation of novel pyrogen tests based on the human fever reaction, with a view to the ultimate replacement of the rabbit pyrogen test and the Limulus assay* (QLK3-1999-00811)

The research project funded by the Commission under the EU Fifth Research Framework Programme (1998-2002) brought together the best teams from academia, industry and regulatory bodies. The Commission's Joint Research Centre (the "ECVAM" facility, or "European Centre for Validation of Alternative Methods") played a major role in the project through provision of scientific and technical advice on the design of the validation study, application of good laboratory practice procedures and distribution and coding of test material.

Industry and regulators jump on board

Interest from both regulatory authorities and industry is very high, with many contributions coming from outside the project consortium that included national control laboratories, test developers, a major pharmaceutical company and a producer of diagnostic kits. For example, the European Pharmacopoeia has set up an international expert group to draft a general method on these new tests. In fact, the tests are already in use in about 200 laboratories worldwide, with great success.

Further take-up and new applications

The Commission will take responsibility for further application of this multidisciplinary, international validation study, including an intended patent. This will encourage successful transfer of the tests and help open new fields for pyrogen testing, such as cellular therapies, medical devices and pollution control in the work place.

Reducing, replacing or refining animal experimentation

Drug quality control is a trans-national matter, which is standardised and regulated in Europe at EU level, thus requiring international collaborative efforts. The European Commission ensures full support for applications to reduce, replace or refine animal experimentation as required by the 1986 Council Directive[2]. This aim is echoed by the European Pharmacopoeia. The "Three Rs" provide a strategy to minimise animal use, without compromising the quality of the scientific work being done.

ECVAM's role is to co-ordinate international validation studies, act as a focal point for the exchange of information, to set up and maintain a database on alternative methods, and to promote dialogue among legislators.

Background: pyrogen and non-oral drugs

Parenteral drugs are commonly employed throughout Europe for treating a variety of illnesses. Ensuring the safety of such widely used drugs requires strict monitoring and control against any possible pyrogenic contamination on a batch-by-batch basis. The most important pyrogen is endotoxin, a constituent of the cell wall of gram-negative bacteria that can generate endogenous fever mediators by white blood cells, particularly monocytes and macrophages.

Rabbits or…

In the rabbit pyrogen test, the test substance is injected into rabbits and any subsequent change in body temperature recorded. A significant rise in temperature indicates the presence of pyrogens. While it has served drug safety control for more than 50 years, it fails for important new therapies such as cellular products or species-specific agents.

[2] *Novel in-vitro testing as alternatives to animal testing*; Council Directive 86/609/EEC

… horseshoe crabs?

Until now, the only in vitro alternative available is the LAL test, based on coagulation of blood from the horseshoe crab (Limulus polyphemus). However the LAL test detects only one class of pyrogens – endotoxins from gram-negative bacteria – leaving patients at risk from "non-endotoxin" pyrogens such as gram-positive toxins, viruses and fungi. It is also subject to interference by various non-pyrogenic substances. And, as it is based on the defence system of an arthropod, it cannot provide results perfectly relevant to humans.

No – human blood cells!

Six alternative cellular assays have therefore been developed to replace the animal rabbit pyrogen test and close the safety gap presented by use of the LAL test in controlling parenterals. All these test systems are based upon the response of human leukocytes (principally monocytes), which release inflammatory mediators (endogenous pyrogens) in response to pyrogenic contamination (exogenous pyrogens).

Quicker, more accurate and more effective

The new tests have several advantages compared with the rabbit test: they are less laborious, cheaper and more sensitive. Results of the validation study suggest that testing on animals can be completely replaced. In contrast to the LAL, the new assays are not restricted to endotoxins from gram-negative bacteria but detect all classes of pyrogens and reflect the potency of different endotoxins in mammals, without suffering interference from endotoxin-binding components in blood products. A commercial kit version for one of the assays has already been developed and standardised, and pre-tested cryopreserved (frozen) blood as a versatile test reagent containing the blood cells as biosensors is under development.

For further information please visit:

http://ecvam.jrc.it/index.htm

http://europa.eu.int/comm/research/quality-of-life/cell-factory/volume1/projects/qlk3-1999-00811_en.html

[This Page Intentionally Left Blank]

Appendix C3

ECVAM Replies to Questions of ICCVAM Pyrogenicity Peer Review Panel

[This Page Intentionally Left Blank]

1 February 2007

ECVAM replies to questions of ICCVAM Pyrogenicity Peer Review Panel

1. Availability of ESAC Peer Review Report

Since we are creating a precedent in making ESAC peer-reviews public, a discussion within ESAC is required, especially since a number of external experts have been involved, who have not been asked. Thus, we are unfortunately not able to make this available at this stage of the process.

2. Lot numbers

a) e-mail of David Allen on 10/01/2007
replied on 12/01/2007 with list of drugs as PDF attached

VALIDATION STUDY: LIST OF DRUGS

Product	Manufacturer	Lot
Alkohol-Konzentrat 95%	B. Braun	2465Z01
Beloc i.v.	Astra Zeneca	DA419A1
Binotal 0,5g	Grünenthal	117EL2
Fenistil	Novartis	21402
MCP Hexal	Hexal	21JX22
Orasthin	Hoechst	W015
Sostril	Glaxo Wellcome	1L585B
Traubenzuckerlösung 5% Eifelfango	Eifelfango	1162

Alkohol-Konzentrat 95% = aethanol 95%
Traubenzuxkerloesung 5% Eifelfango = 5% glucose solution

b) e-mail of David Allen on 12/01/2007
Yes, individual lots were tested in all methods during the validation/catch-up validation study

However, some of the lots used in the validation study were no longer available for the catch-up validation study and one product (Orasthin) was no longer on the market. It was replaced with a product (Syntocinon) containing the same active ingredient. Please find attached the pdf file "List of drugs catch-up validation" and the table below highlighting differences in lot numbers and products.

VALIDATION STUDY (CATCH UP): LIST OF DRUGS

Product	Manufacturer	Lot
Alkohol-Konzentrat 95%	B. Braun	2465Z01
Beloc i.v.	Astra Zeneca	DA419A1
Binotal 0,5g	Grünenthal	117EL2
Fenistil	Novartis	26803
MCP Hexal	Hexal	21JX22
Orasthin	Hoechst	not available
Sostril	Glaxo Wellcome	3H01N
Syntocinon 3 I.E.	Novartis	S00400
Traubenzuckerlösung 5% Eifelfango	Eifelfango	3132

Alkohol-Konzentrat 95% = aethanol 95%
Traubenzuxkerloesung 5% Eifelfango = 5% glucose solution

3. GLP concordance

e-mails of David Allen on 9/01/2007 and on 12/01/2007 (question 1)

a) In vitro data

The initial validation study has been carried out to large extent in laboratories such as National Control laboratories, which do not operate under GLP. It was, however, agreed to comply with the requirements of GLP, especially with regard to the creation and management of SOPs. The partner laboratories have received presentations on the requirements. No auditing was done but various quality checks and blinding mainly under the responsibility of ECVAM were included.
In the catch-up validation, two GLP laboratories and two National Control laboratories participated.

Raw data: In both studies the laboratories were asked to transfer the readings into the excel sheets provided by the biostatistician. This was mostly done by directly inserting the ASCII files created by the plate reader. However, reader printouts are available and can be provided on request.

b) In vivo data and reference to Section 4.4 in the ECVAM BRDs
Indeed it should read here "not applicable" as stated in the WB/IL-6 BRD, since the RBT was not performed during the validation study. As indicated in 4.1 the data used were provided by the Paul-Ehrlich-Institute (PEI; www.pei.de), which is the German Federal Agency for Sera and Vaccines (competent authority) and conducts the RBT according to the European Pharmacopoeia. For further information on the quality assurance established at the PEI please contact Dr Thomas Montag (e-mail: month@pei.de).

This should also be corrected in the main document 4.4 In vivo data quality.

4. Data analysis
e-mail of David Allen 12/01/2007 – question 2

The same data analysis was applied in both studies. The first paragraph in Section 5.3 reads *A generally applicable analytical procedure was employed. This procedure includes a universal PM as well as quality criteria. First, a two-step procedure consisting of a variance-criterion and an outlier-test was applied. For this, the Dixon's test (Barnett and Lewis, 1984), which is USP approved, was chosen with the significance level of* α*=0.01 and applied to identify and eliminate aberrant data.*
Please find attached to our mail, the document 'Trial data report' of the validation study. It was not included in the submission to ICCVAM, since a lot of the procedures described here are included in the BRD. Related to your question, you will find in Chapter 4.2 the procedure describing the exclusion of data. There, the Coefficient of Variation (CV) was used as a trigger to investigate the replicates of a given control or sample. Excessive variability would severely impair the prediction model, resulting mainly in a loss of specificity. The CVs were empirically determined for each assay based on the information collected in the protocol optimization phases (Phase A & B) and the prevalidation. Thus, they can differ between assays.
In addition, we attach the document *Analytical procedure to identify and eliminate outlying observations* written by the responsible statistician, Sebastian Hoffmann, during the validation study and which gives rationales for applying this procedure.

5. Selection of test substances
e-mail of David Allen 12/01/2007 – question 3

Please find attached the file "Rationale for selection of test substances".

6. Removal of DMSO
e-mail of David Allen 12/01/2007 – question 4

Schindler et al 2004 state:

We sought to develop a protocol which would allow the use of the thawed whole blood samples directly without any washing steps to remove the cryoprotectant, as such a step would eliminate the essential advantages of the human whole blood assay, i.e., the ease of performance which allows a high degree of standardization as shown for various applications (Fennrich et al., 1999). Furthermore, besides stress and handling artifacts, the cells would lose their autologous plasma that permits a number of physiological responses, e.g., the sensitive response to

lipopolysaccharides (endotoxin, LPS) via lipopolysaccharide binding protein (LBP; Schumann, 1992; Fenton and Golenbock, 1998).

Indeed DMSO is not removed and up to now artefacts attributed to the presence of DMSO were not observed. The presence of DMSO enhances the IL-1 production and leads to a delay in the release. The fact that no wash step is required reduces strongly variation and introduction of artefacts.

7. Possible cytotoxicity
e-mail of David Allen 12/01/2007 – question 5

The aspect of cytotoxicity is covered by interference testing. As stated already on various occasions, interference testing (what we called positive product control in the validation study) is a must before you can use the WB (and the other) assays. If a substance would interfere with the assay by being cytotoxic, the spike recovery would be below 50%.

8. Freeze-thaw step for CRYO WB/IL-1 (Konstanz method)
e-mail of David Allen 12/01/2007 – question 6

This question was already posed during the drafting of the ICCVAM peer review documents (Mail David Allen 1/08/2006 question 3 and attached document PyroProtocol31Jul06)

In our reply (sent on 8/09/2006 with attachment reply_PryoProtocol31Jul06), we stated on page 2: *The freezing thawing enhances the IL-1 release and makes the Konstanz method more robust and reliable. It is not needed for the PEI method since the IL-1 release levels are higher.*
In fact, it should read that the freezing thawing enhances the IL-1 yield since the IL-1 produced in but not released by the monocytes is also measured.
It has been shown by Boneberg and Hartung (2003) that 10fold higher concentrations of (pro-)IL-1ß are found when including intracellular cytokine by whole blood lysis:

Ref *Boneberg E. and Hartung T. Febrile temperatures attenuate IL-1β release by inhibiting proteolytic processing of the proform and influence Th1/Th2 balance favoring Th2 cytokines. J. Immunol. 2003, 171:664-8.* attached.

Appendix C4

Rationale for the Selection of the 10 Substances Tested in the Validation/Catch-Up Validation Study of *In Vitro* Assays for Pyrogen Testing

[This Page Intentionally Left Blank]

Rationale for selection of the 10 substances tested in the validation/catch-up validation study of in vitro assays for pyrogenicity testing

Selection committee:
Thomas Montag-Lessing (chair), Michael Jahnke, Ingeborg Aarberge, Sandra Coecke

The main points which led to the selection were stability of the spikes, relevance, availability and costs of the substances:

1. Stability of the spikes, coding, interference testing

Experiments to evaluate the stability of endotoxin spikes in the final products revealed that stability of low endotoxin concentration could not be guaranteed over the time period needed for the prevalidation/validation study. Therefore, endotoxin spikes in higher (stable) concentrations were produced, filled in separate vials and coded. The laboratories received the clean substance plus the coded spikes, the clean substance had to be used for interference testing and contaminated with the coded spikes for the actual tests.

2. Relevance

The absence of pyrogens is crucial for intravenously administered drugs, this is reflected in the rabbit pyrogen test where the test substance is injected into the ear vene.
Therefore, only substances intended for i.v. injection were selected. In addition, it could be evaluated whether the in vitro assays would be able to detect 0.5 IU/ml endotoxin, which corresponds to threshold inducing fever in rabbits.

3. Availability/feasibility

- Substances should be on the market, thus the final product in the original vials could be tested and the conditions under which a lab performing final lot release would work could be met, e.g.
    - avoid possible contamination with pyrogens during opening the vials, drawing the samples etc
    - performing interference testing (corresponds to positive product testing in the validation trial)

- One lyophilised product was included in order to check for potential failures (e.g. pyrogenic contamination during reconstitution of the drug)

- Substance not interfering with any of the assays in order to control the correctness of the spiking procedure. Therefore, 0.9% NaCl pyrogen-free solution was included (Drug A and B)

4. Costs

- Due to the restricted funds available, costs of the substances to be tested in the validation trial played a role, e.g. it was not possible to include a rather expensive blood product as coagulation Factor VIII (Haemate® was used in the prevalidation trial).

Appendix C5

Comparison and Validation of Novel Pyrogen Tests Based on the Human Fever Reaction: Trial Data Report

[This Page Intentionally Left Blank]

Validation of Biomedical Testing Methods

Comparison And Validation Of Novel Pyrogen Tests Based On The Human Fever Reaction

Acronym: Human (e) Pyrogen Test

Trial data report

Date of Circulation:
File Identifier: *Stp-HPTVv03*

JOINT RESEARCH CENTRE
EUROPEAN COMMISSION
Institute for Health and Consumer Protection
ECVAM

TABLE OF CONTENTS Page no.

Comparison And Validation Of Novel Pyrogen Tests Based On The Human Fever Reaction

Acronym: Human (e) Pyrogen Test

1 INTRODUCTION

The objective of the "Human(e) Pyrogen Test" project is to assess the performance and use of six recently developed *in vitro* pyrogen tests. These tests are based on the human fever reaction. As they are meant to be similar to the currently used Rabbit Test, the analytical procedure is designed to give a dichotomous outcome. In detail, drugs, which have to be tested for pyrogenic contamination due to regulatory instruction, have to be classified either as hazardous, i.e. pyrogenic, or as safe for humans. Hence, securing the safety of humans is the primary objective of pyrogenicity testing in general. Therefore, the prediction model is constructed to give a clear-cut classification of a given drug taking the safety aspect into account.

In addition to the prediction model, procedures to ensure quality criteria the test systems have to meet are included. A two-step method to identify and eliminate aberrant data as well as a test for a sufficient limit of detection are provided. Information from previous phases of the project established the basis to develop and define these methods. The data from the pre-validation study were used to refine the procedures.

All methods of the analytical procedure were developed to be applicable to each of the six test systems and were accepted by the participants.

.

2 THE BASIC BIOLOGICAL PRINCIPLE

The six test systems make use of the same biological principle. The mediators of the human fever reaction are cytokines, which are produced by monocytes. This principle is employed by incubating either fresh human cells or cell lines with the drug to be analysed under SOP-defined conditions. As there are several cytokines, which highly correlate with the human fever reaction, the cytokine of choice of the test sytems differs. Mainly the cytokine IL-6, but also IL-1β, TNF and neopterin were chosen as endpoints. After the incubation, an also SOP-defined ELISA-step is performed. In this step the cytokine is bound, visualised and finally measured by an optical reader. As the visualisation of the endpoint, measured as optical density (OD), is proportional to the amount of cytokine present, the resulting data are metrically scaled. In these entire procedure, a biological standard, WHO-LPS 94/580, is employed as an objective tool for comparison.

One of the test systems is based on a competitive ELISA, which results in a monotonically decreasing dose-response relationship, whereas the other systems show an increasing relationship due to their sandwich ELISA technique.

3 STATISTICAL PROPERTIES OF THE DATA

In previous phases the statistical properties of the data were analysed. Replicate observation for a fixed known control or an unknown drug revealed a right-skewed distribution. In experiments with large sample sizes it was shown, that a ln-transformation of the raw OD-data allows to assume a gaussian distribution of the data, which parameters can be estimated by the mean and the empirical variance.

As handling errors in the conduction of the test result in extreme observations, which may have an crucial impact on the prediction model, the probability of occurrence and impact of these observation was analysed. Although the probability of extreme observations is small for all tests, the inclusion of a method to identify and eliminate these data is indicated to ensure an optimised performance of the prediction model.

Furthermore, it was confirmed, that the dose-response relation ship between concentration of the contamination and the response increases, respectively decreases, monotonically for increasing concentration.

4 METHODS

4.1 Background

The entire analytical procedure consists of three different techniques, two of which assure the appropriateness of the data. The ELISA-plates employed have a 96-well format. The data of one plate have to be considered as a whole, which can not be compared to other plates due to uncontrollable variation. Therefore, each plate has to include all controls required for the analytical procedure. These are a negative control, which is 0.9%-NaCl, and a positive control of the WHO-LPS 94/580 standard diluted in 0.9%-NaCl, as well as negative and positive controls of the drugs, which are to be tested on the plate. Negative controls of a drug are obtained by released batches of the drugs. Positive controls are gained by adding 0.5 endotoxin units (EU)/ml of WHO-LPS 94/580. These 0.5 EU/ml were concordantly defined as the threshold concentration of endotoxin that induces fever in humans under worst conditions. In previous experiments it was shown, that this positive control lies in the most sensitive region, i.e. the steepest part, of the dose-response curve of all six test systems. In the following the NaCl-controls are denoted as "C-" (negative) and "C+" (positive). Similarly, the controls of a drug S_i are denoted as "S_i-" ans "S_i+". Furthermore, "S_ij", j=1, 2, 3, 4, 5, represent the blinded versions of the drug S_i.

4.2 Method A: Identification and elimination of aberrant data

The first method to be applied is an method to check the quality of the data of a plate. In general, this is done by a two-step procedure, which firstly identifies the sets of replicates with an extremely large variation. A set of replicates consists of four replicates per control, respectively drug tested. For every test system a maximal coefficient of variation (CV_{max}) was extracted from the available information. If the CV of a set of replicates is smaller than its CV_{max}, it is analysed as it is. Otherwise, the set is examined in the second step. This second step is a test for outliers. Therefore, the Dixon's test (1), which is USP approved, was chosen with the significance level of α=0.01. Preliminary to the testing itself, the raw OD-data are transformed with the natural logarithm, which normalises the data to meet the

prerequisites of the Dixon's test. If one observation in a set, which is identified by the Dixon's test, is responsible for its large variation, then this observation is excluded. If the variation is due to all observations, i.e. the absence of an outlier, the entire set of replicates is excluded from further analysis. Unfortunately, this approach poses the danger, that a whole plate can not be analysed, when a control is to be excluded. Therefore, both steps were chosen conservatively. Nevertheless, the empirical nature of the first step is not optimal and depends on general properties of the test system. But all established statistical methods, which address this problem, e.g. the Bartlett test for heterogeneity of variances, are not appropriate, because the variance structure over the range of concentration is highly variable and their global character. In table 1 the empirically derived CV_{max} are listed for the six test systems. The approach could be harmonized over all test systems.

test system	MM6	PBMC	THP-1BN	THP-1IK	WBT-KN	WBT-NI
CV_{max}	0.25	0.45	0.45	0.25	0.45	0.45

Table 1: Maximum CV's for the six test systems

4.3 Method B: Assuring the limit of detection

The second method is designed to ensure an minimum limit of detection of a plate (2). Because of the pre-defined dichotomous classification, a crude criterion, which merely shows strict monotonicity in the interesting part of the dose-response curve, can be chosen. Therefore, a one-sided t-test with a significance level of α=0.01 is applied to the ln-transformed data to ensure, that the response to the positive control is significantly larger than that of the respective negative control.

4.4 Method C: The prediction model

The third and most important statistical tool is the so-called prediction model (PM). In general, it is a statistical model, which classifies a given drug by an objective diagnostic or deciding rule. The objective of a dichotomous result requires a clear cut PM, which assigns a drug in one of the two classes "pyrogenic for humans" and "non-pyrogenic for humans". As the members of the project decided on a threshold positive

control, a one-sided test is appropriate for the task. Because the data are normalised by a ln-transformation, a t-test was chosen. Although the variances over the range of concentration converge by the transformation, the assumptions of equal variances does generally not hold true, because it depends on additional covariates. Therefore, the one sided Welch-t-test (3) is applied. Due to the safety aspect of the basic problem, the hypotheses of the test are

$$H_0 : \mu_{S_i j} > \mu_{S+} \quad vs \quad H_1 : \mu_{S_i j} < \mu_{S+},$$

where $\mu_{...}$ denotes the parameter of location of the respective ln-transformed distribution. This approach controls the probability of false positive outcomes directly by means of its significance level α, which is chosen as 0.01, because is assumes hazard, respectively pyrogenicity, of the tested drug in H_0, and assures safety, i.e. non-pyrogenicity. The test statistic is

$$T_{S_i j} = \frac{\bar{x}_{S+} - \bar{x}_{S_i j}}{\sqrt{\frac{s^2_{S+}}{n_{S+}} + \frac{s^2_{S_i j}}{n_{S_i j}}}}.$$

The PM is built by means of the outcome of the test. Let 0 denote safety and 1 denote hazard. The classification of S_i-j is then determined by

$$S_i j = 0, \text{ if } \quad T_{S_i j} > t_{0.99; n_{S+} + n_{S_i j} - 2},$$

$$S_i j = 1, \text{ else,}$$

where $t_{0.99; n_{S+} + n_{S_i j} - 2}$ the 0.99-quantile of the t-distribution with $n_{S+} + n_{S_i j} - 2$ degrees of freedom. The number of replicates for every control and sample, i.e. $n_{...}$, was harmonised for all test systems to be four. Due to the possibility of removing one observation by the outlier test, the number of replicates could be reduced to three. The classification of a version of a drug is regarded as an independent decision. Therefore, the niveau α is local.

4.5 Method D: 2x2 contingency tables for the final results

Finally, the classifications of the drugs will be summarised in 2x2 contingency tables, formally presented in table 2.

		pre-defined class		Σ
		1	**0**	
classification by test system and PM	**1**	**a**	**b**	**a+b = $n_{.1}$**
	0	**c**	**d**	**c+d = $n_{.0}$**
Σ		**a+c = $n_{1.}$**	**b+d = $n_{0.}$**	**n**

Table 2: 2x2 contingency table

From these tables estimates of the sensitivity, i.e. the probability of correctly classified positive drugs, and specificity, i.e. the probability of correctly classified negative drugs, will be obtained by the respective proportions. Furthermore, these estimates will be accompanied by confidence intervals, which will be calculated by the Pearson-Clopper method (4). For example, let $\hat{p}_{SE}$ denote the proportion, namely the sensitivity, under investigation. Then the confidence interval to a niveau α is calculated as

$$\left[p_{SE}^{L} = \frac{\mathrm{a} F_{2\mathrm{a};2(n_{1.}-\mathrm{a}+1);\frac{\alpha}{2}}}{n_{1.}-\mathrm{a}+1+\mathrm{a}F_{2\mathrm{a};2(n_{1.}-\mathrm{a}+1);\frac{\alpha}{2}}}; p_{SE}^{U} = \frac{(\mathrm{a}+1)F_{2(\mathrm{a}+1);2(n_{1.}-\mathrm{a});1-\frac{\alpha}{2}}}{n_{1.}-\mathrm{a}+(\mathrm{a}+1)F_{2(\mathrm{a}+1);2(n_{1.}-\mathrm{a});1-\frac{\alpha}{2}}} \right],$$

where $F_{\ldots}$ denotes the respective quantile of the F-distribution and $n_{1.}$ is the sample size of the positive drugs and a the number of correctly classified drugs. By contaminating the drugs artificially and by defining a threshold dose, which is assumed to be appropriate, the class of a drug is determined beforehand. The versions of drugs, which are effectively contaminated, but below the threshold dose, are considered to be negative, respectively safe, because their contamination is not crucial for humans in terms of endotoxin limit concentration.

5 ANALYTICAL PROCEDURE

5.1 General procedure

The process of the analytical procedure is highlighted in figure 1. Firstly, the data of the controls of a ELISA-plate are checked for aberrant data with procedure A. If indicated, outliers are removed. If sets of replicates are to be removed, this is recorded, but due to the empirical base of the first step of A, the data will be further analysed with reservations. Afterwards, the remaining data of the controls are tested with method B to ensure a minimum limit of detection. If the controls in 0.9%-NaCl do not differ significantly, the further analysis is done with reservations. If the controls of a drug do not differ significantly, all data of this drug do not qualify for further analysis. The last part of quality assurance is the application of method A to the data of the blinded drugs. Here, drugs, which fail the criteria, are removed from further analysis. Finally, the remaining data are put to the prediction model. The classification of the still blinded drugs are sent to ECVAM in an official document, which will in general comprise the assigned class for every drug structured by test system, laboratory and drug. Upon receipt ECVAM will send the blinding code in a electronically generated document by e-mail to the project's statistician Sebastian Hoffmann (e-mail: sebastian.hoffmann@uni-konstanz.de). Additionally, a hardcopy of the blinding code will be sent by post. Once the data are unblinded, the final results, which core will be method D, can be summarised, explicitly analysed and appropriately presented. Additionally to the contingency tables and related topics, an inter-laboratory comparison will be done. Furthermore, the reasons for misclassifications will be identified.

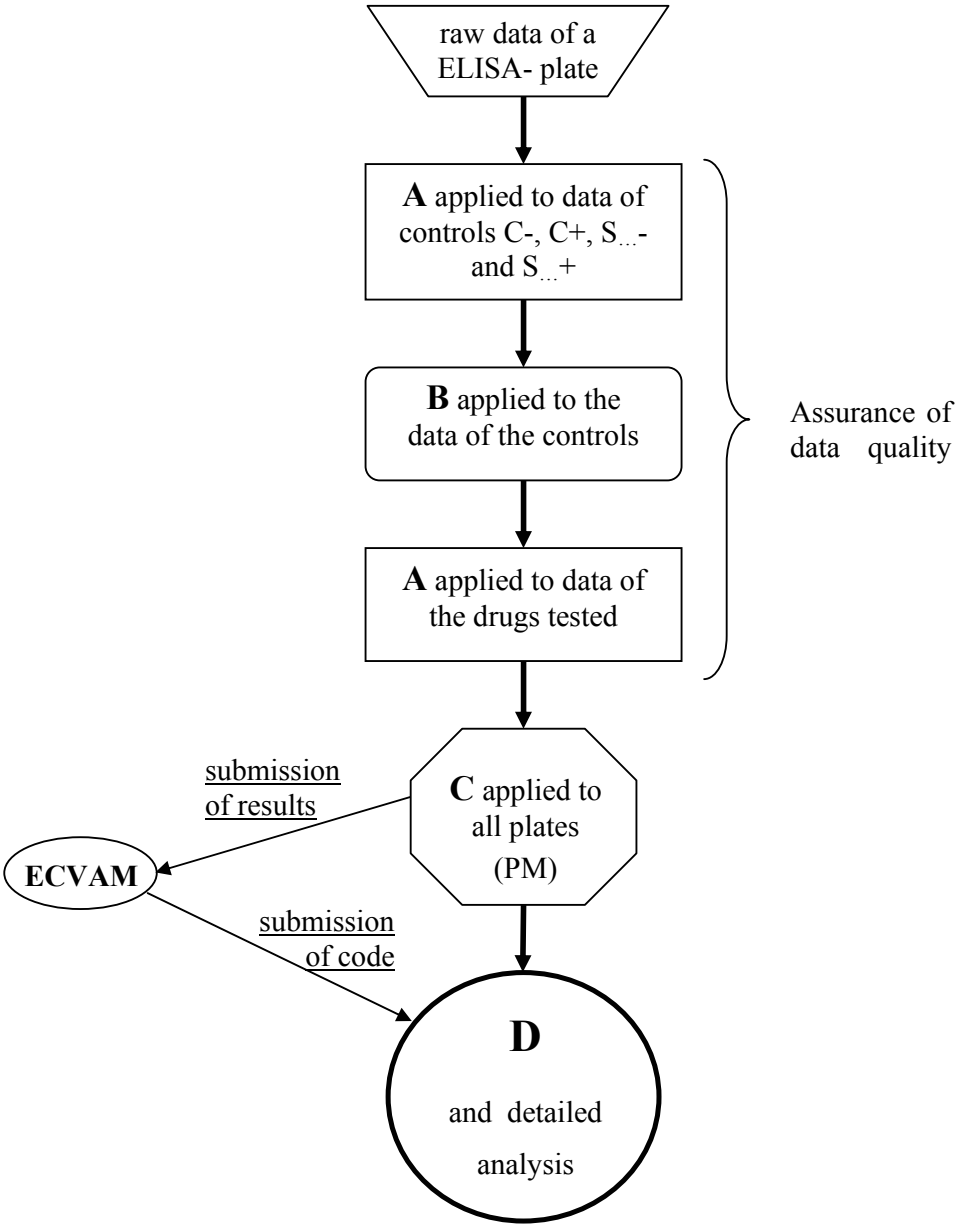

Figure 1: Flowchart of the main analytical procedure

5.2 Modification for the test systems PBMC and WBT-NI

This analytical procedure has to be modified for the two test systems PBMC and WBT-NI. The two test systems base their classification of a drug not on one outcome, but they classify a drug by the results of several independent experiments. Because both methods rely on fresh blood, the inter-donor variability is taken into account by using the blood of several donors and conducting the test independently. The modifications, which have to be made, arise out of contradictory classification of a drug by different donors. Therefore the classification of a drug is determined by the combination of the single donor-dependent results, which are calculated with the analytical procedure presented.

6 ADDITIONAL ANALYSIS

The detailed analysis will contain an inter-lab comparison per test system, whereas measures of correlation as well as similarity can be used. Furthermore, shortcomings depending on specific drugs, laboratories, the analytical procedure and/or test systems will be examined by exploratory statistical methods.

Additionally, modifications of the methods A, B and C will be used to optimise the analytical procedure with the information from the new data. In general, these are the consequences of more restrictive or less restrictive assumptions. With regard to A, the results of a procedure without a tool for aberrant data will be compared to the results of the described procedure. For B, more restrictive criteria to ensure a valid dose-response relationship will be applied, e.g. techniques for ratios between controls based on Fieller's theorem (5, 6). Besides, modifications in the t-test of the prediction model will be of interest, mainly assumptions considering the variance and a multiple testing approach. E.g. a simulation, which allows for the $\sqrt{k}$-rule optimising the Dunnett's test could be realisable.

Finally, methods taking the real life situation of pyrogen testing into account will be highlighted. These include a Fieller-based method to handle interference.

7 REFERENCES

1. Barnett, V., Lewis, T. (1984). *Outliers in statistical data*. 2nd edition, Chichester, US: John Wiley & Sons, pp 171-172.
2. Hothorn, L. A. (1995). Biostatistical analysis of the 'control vs. k treatments' design including a positive control group. In *Testing principls in clinical and preclinical trials*, pp 19-26, Stuttgart, Germany: Gustav Fischer Verlag.
3. Snedecor, G. W., Cochran, W. G. (1989). *Statistical methods*, 8th edition, Ames, US: The Iowa State University Press.
4. Clopper, C. J. & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404-413
5. Fieller E. (1954) Some problems in interval estimation. *J. Royal Statistical Society* **B16**, 175-185.
6. Hauschke, D., Hothorn, T., Schaefer, J. (2002). The role of control groups in mutagenicity studies: matching biological and statistical relevance. *ATLA* (in press)

Appendix C6

List of Drugs for the Catch-Up Validation Study

[This Page Intentionally Left Blank]

VALIDATION STUDY (CATCH UP): LIST OF DRUGS

Product	Manufacturer	Lot
Alkohol-Konzentrat 95%[1]	B. Braun	2465Z01
Beloc i.v.	Astra Zeneca	DA419A1
Binotal 0,5g	Grünenthal	117EL2
Fenistil	Novartis	26803[2]
MCP Hexal	Hexal	21JX22
Orasthin[3]	Hoechst	not available
Sostril	Glaxo Wellcome	3H01N[4]
Syntocinon 3 I.E.	Novartis	S00400
Traubenzuckerlösung 5% Eifelfango[5]	Eifelfango	3132[6]

16.02.2004

[1] "95% Alcohol Concentration"
[2] Different lot number
[3] Orasthin no longer on the market, replaced with Syntocin 3 I.E. containing also oxytocin
[4] Different lot number
[5] "5% Glucose Solution"
[6] Different lot number

[This Page Intentionally Left Blank]

Appendix C7

Analytical Procedure to Identify and Eliminate Outlying Observations

[This Page Intentionally Left Blank]

Analytical procedure to identify and eliminate outlying observations

Introduction

As seen in the pre-validation, the problem of outlying observations is not appropriately solved yet. The crucial issue about these observations is their impact on the prediction model, which could result in false classifications of substances. Of course, one could just neglect such observation, as done in the pre-validation. This is the most easy way, but its appropriateness with regard to sensitivity and specificity is depending on the probability of outlying observations. So far, this probability was low, smaller than 5%, for all assays. Nevertheless, even if there are only a few outliers, this approach has the disadvantage, that one would have to live the most extreme and maybe even pre-identified outlying observations, e.g. when a technician recognises that she/he made a gross handling error.

Therefore, a new analytical procedure was developed. First of all, the objective of such an procedure has to be defined precisely. On the one hand, a way to identify obvious handling errors, which most often can be identified by eye by trained persons, is needed. On the other hand, a method to handle sets of replicates, which are extremely untypical for a specific assay, has to be taken into account. Hence, a generally applicable two-step procedure is proposed.

Step 1: Checking the variation between sets of replicates

Firstly, the data of an ELISA-plate are checked for untypical variation of one or more sets of replicates. In the given situation, one is only interested in those sets with extremely large variance. In general, there are two situations which have to be considered. On the one hand, just one observation could be responsible for a huge variance in its set of replicates. On the other hand, equally distributed replicates over a large range of response, which includes the situation of two outliers in a set, might be the reason.

The existing statistical tests addressing this question, e.g. the Bartlett-test, are not appropriate due to various reasons, but mainly because they assume homogeneity of variances and show global heterogeneity. Thus a simple empirical method was derived for every assay, which is mainly based on the data of the pre-validation and the information from Phases A and B. The core of this method is an appropriate measure of variation for a set of replicates. Here, the CV is chosen, but the variance or the standard deviation can be used more or less equivalently. From these empirical information, a maximum CV, denoted as CV_{max}, was derived, which can be used as a tool to assess the variation of each set of replicates very easily. If a CV of a given set is larger than CV_{max}, then this set will be examined further in the second step of the

procedure. If the CV is smaller, then the data of the set will be analysed as they are. To highlight this empirical method, it is exemplarily explained for the Novartis-PBMC assay. In figure 1 the variation within sets of replicates for all available data is presented.

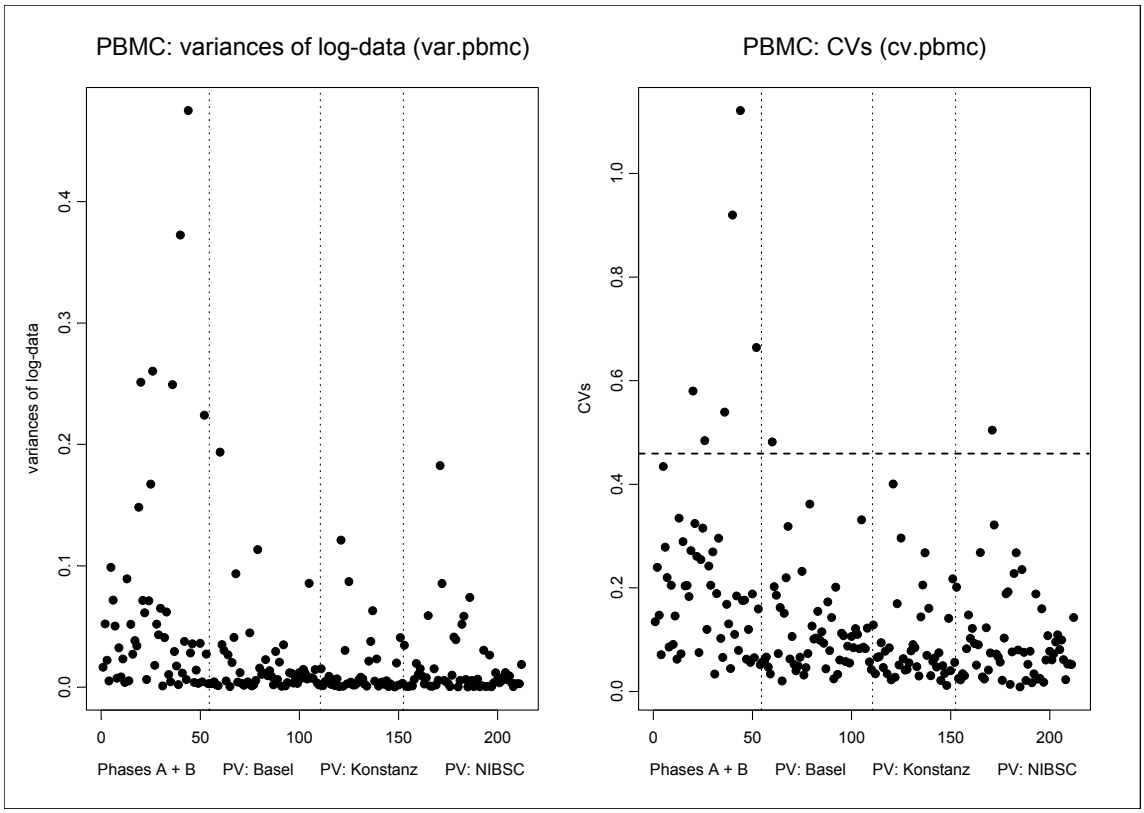

Figure 1: Variation within replicates for the PBMC-assay

Together with the raw data and some linear modelling techniques, here $CV_{max}(PBMC) = 0.45$ was chosen. This choice identifies two out of 154 pre-validation data sets as outlying sets in the right part of figure 1. As can be seen in the left part of figure 1, a criterion based on the variance, e.g. $Var_{max}(PBMC) = 0.18$ is almost identical, which even can be shown by some statistical approximation under certain assumptions.

Additionally to the approach with the CV, a criterion based on the ratio of variances was applied. Also having the empirical background, it did not show any advantage.

Step 2: Checking the variation within sets of replicates identified in step 1

Let S_i, i=1,…, denote the crucial, in step 1 identified sets with $CV(i) > CV_{max}$. In this second step the reasons for the high variation of the S_i's are examined. Firstly, a common test for outliers, the FDA-approved Dixon-test, is applied to each S_i with the niveau α of 1%. If an outlier is identified, it is withdrawn from its set and the remaining data are further analysed. If

no outlier is detected, the observations of a set of replicates are regularly distributed over a large range of response. In the latter case, it is recommended, to repeat the substance(s) S_i on another plate.

Discussion

Assuming such a partly empirical approach is appropriate, one still has to be aware of its properties and effects, especially when applied in the validation study. Firstly, the procedure gives excellent results when applied to the pre-validation data. This is expected, because the CV_{max} criterion was mainly derived by the data themselves, which makes it a self-fulfilling prophecy. Therefore, it poses the danger of choosing the CV_{max} too small, because it may lead to a lot of rejected sets. Additionally, maintaining such an empirical procedure demands to check regularly for the validity of the chosen $CV_{max.}$

Secondly, in the given situation of the validation study, the impact of the retrospectively applied procedure has to be taken into account. Because the data are checked outside the labs, the sample size of the number of classified samples could be reduced during analysis. For example, assume that a control on a plate, on which the prediction model is based (e.g. the positive product control), does fail the above proposed procedure. In the case of the positive product control, this would mean that none of the samples tested on that plate could be classified by the prediction model.

Furthermore, the robustness of the procedure with regard to systemic errors is noteworthy. It will work, even if the ELISA-plate is of low quality, e.g. with regard to coating, or if moderate systemic handling errors are present.

Application of the procedure to the available data

assay	CV_{max}	number of sets	Outlier procedure		Dixon test
			outliers	outlying sets	
THP-Bern	0.45	138	1	4	7
THP-Inns.	0.2	112	-	-	3
MM6	0.2	129	1	1	5
PBMC	0.45	154	1	1	6
WBT-Konst.	0.45	138	-	1	3

Table 1: Results of the outlier procedure

The results in table 1 are very promising, but should not be overestimated as mentioned above. For example, the problematic first run from Oslo with the THP-Bern assay with regard to Haemate can easily be handled with the proposed procedure, because three of the Haemates would have to be retested. In contrast, the Dixon test alone would not have detected any outlier in the three Haemates.

As can be seen in table 1, a harmonised choice for CV_{max} was sought. Alternatively, the more conservative $CV_{max} = 0.25$ for the two-plate cell line assays (THP-Innsbruck, MM6) could have been applied giving very similar results. The more restrictive $CV_{max} = 0.4$ for the THP-Bern and the two methods based on fresh blood could have also been chosen.

Unfortunately, the variation within sets of replicates for the WBT-NIBSC increased from Phases A and B to the pre-validation and is fortunately decreasing at the moment due to changes in the SOP. But considering the variation shown in Phases A and B and the harmonising aspect of the above proposal, a CV_{max}(WBT-NIBSC) = 0.45 seems to be appropriate.